AD_____

Award Number: W8XWH-05-1-0236

TITLE: A Novel Approach for the Identification of Pharmacophores through Differential Toxicity Analysis of Estrogen Receptor Positive and Negative Cell Lines

PRINCIPAL INVESTIGATOR: Albert R. Cunningham, Ph.D.

CONTRACTING ORGANIZATION: University of Louisville Research Foundation
Louisville, KY 40014

REPORT DATE: July 2009

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
07/31/2009	Annual	1 Jul 2008 To 30 Jun 2009

4. TITLE AND SUBTITLE; A novel approach for the identification of pharmacophores through differential toxicity analysis of estrogen receptor positive and negative cell lines

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W8XWH-05-1-0236

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Albert R. Cunningham, Ph.D.

Email: al.cunningham@louisville.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Louisville Research Foundation
Louisville, KY 40014

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research
Research and Materiel Command
Fort Detrick, Maryland
21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This grant was awarded to the PI at Louisiana State University. He left there in Aug 2006 and with Dr. Carole Christian, it was transferred to the University of Louisville. Minimal work was done on the project prior to departure from LSU and minimal funds were expended. The PI's new laboratory has been equipped with the needed computer hardware and software and a postdoctoral fellow (C. Alex Carrasquer) and research associate (Shahid Qamar) work on this project. Briefly, Aim 1 is essentially completed and several publications are in preparation, results for Aim 2 have also been generated, and in conjunction with Aims 1 and 2, a list of compounds for in vitro testing in Aim 3 have been forwarded to the project's collaborator Dr. Billy Day at the University of Pittsburgh. Dr. Day is currently working on getting physical samples from the NCI for testing, has established the protocols for his aim, has published several manuscripts related to this project, and his graduate student Miranda Sarachine, partially funded by this project, has graduated with a Ph.D.

15. SUBJECT TERMS
Structure-activity relationship modeling, proteomics, drug discovery

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	14	**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U			

Table of Contents

Page

Introduction

The objective of this project is to employ an innovative approach to discover new molecular targets found only in estrogen receptor positive (ER+) breast cancer cells that are 1) highly associated with cell type-specific toxicity, 2) compounds that influence or interact with them, and 3) ultimately anticancer pharmacophores that uniquely target breast tumor cells to be used as the basis for the design of new anticancer drugs. We observed that certain chemicals display potent toxicity to one type of breast cancer cell line and not other related lines. Based on this observation, the **hypothesis** for the project is that this cell type-specific toxicity is due to an interaction of a chemical agent with a specific molecular target found only in the sensitive cells. The project is defined by two working hypotheses. The **first working hypothesis** is that congeneric sets of compounds that display this excessive and specific toxicity to ER+ cells will influence particular molecular targets found only within the sensitive cell line. The reasoning for this is based on the accepted premise of structure-activity relationship (SAR) modeling that like structure begets like activity. The **second working hypothesis** is that, when the proteome of the ER+ cell line is probed with a defined congeneric series of compounds that display this cell type-specific activity, these compounds will all affect, minimally, the same identifiable molecular target. Through the techniques of comparative proteomics, we anticipate being able to identify these unique target(s) and thus provide the basis for highly effective antibreast cancer therapies.

Body

As mentioned in the previous Annual Report, this grant was awarded to the PI at Louisiana State University. The PI left there in Aug 2006 and with the assistance of Grant Managers Drs. Christian and Fallas, the grant was transferred to the University of Louisville where the PI started Apr 2007. Minimal work was done on the project prior to departure from LSU and minimal funds were expended. This project was essential to the PI obtaining an appointment as an Associate Professor of Medicine with a joint appointment as Associate Professor of Pharmacology and Toxicology at the University of Louisville's James Graham Brown Cancer Center. Furthermore, this project led to a significant startup package from the University of Louisville and involvement as a Project PI on the Brown Cancer Center's NIH-funded Molecular Targets Program.

The PI's new laboratory was equipped with the needed computer hardware and software described in the proposal. In August 2007, Shahid Qamar was hired to fill the postdoctoral fellow position. At the time, Shahid was nearing the completion of his Ph.D. studies at Arizona State University and has since graduated. Additionally, C. Alex Carrasquer joined the lab as a post-doctoral fellow. Currently, Drs. Carrasquer and Qamar are each 50% effort on this project.

The specific aims for the project are:

1. Model Development: Create high-quality "control" SAR models for breast cancer cell lines based on 50% growth inhibition, total growth inhibition, and 50% lethal concentration (GI_{50}, TGI, and LC_{50}) values and "experimental" excessive toxicity models based on compounds that display potent and specific toxicity to ER+ cells and minimal toxicity to other comparable cell types.

2. Pharmacophore Identification: Based on structural information developed in Aim 1, identify congeneric sets of compounds associated with excessive and specific toxicity to ER+ cell lines. Define and fine-tune pharmacophores from these models with ligand-based three dimensional SAR methods. Based on these pharmacophores, develop small libraries of available and suitable compounds for in vivo testing. Verify the association between the pharmacophore and defined cell toxicity in vivo with selected compounds.

3. Proteome Target Identification: Using protein mass spectrometric techniques (i.e., proteomics), identify the specific molecular targets associated with excessive and specific toxicity to ER+ cells that are influenced by these pharmacophore-defined sets of compounds and ascertain their involvement in ER+ -cell type-specific toxicity.

Specific Aim 1:

As described in the proposal, the project is based on the observation that certain chemicals display potent toxicity to one type of breast cancer cell line and minimal or no toxicity to other related ones. By comparing chemicals that exhibit this "excess" or "differential" toxicity to one line and not another, the project's goal is to identify what is interesting or critical about these chemicals that gives rise to cell type-specific toxicity. Our criteria for chemical identification and subsequent modeling and target identification is greater toxicity toward estrogen receptor (ER) positive (ER+) cell lines than other similar cell lines. The project started with comparisons between ER+ MCF-7 breast cancer tumor cells and ER negative (ER-) MDA-MB-231 ones and vice versa. (This modeling approach has since been extended to several other breast and non-breast cancer cell lines discussed below).

This original contrast between MCF-7 and MDA-MB-231 cells may seem to overly simplify the problem and suggest that we may rediscover the ER. However, absence or presence of the ER is only one attribute separating MCF-7 from MDA-MB-231 cells. We know that the ER and its signaling pathway are associated with many components including regulators, cofactors, metabolic enzymes, and transport mechanisms. Thus the method allows for the identification of agent-target interactions at numerous points associated with ER status.

With the success of the MCF-7 and MDA-MB-231 models, we extended the analysis to include other breast cancer cell lines (i.e., BT549, HS578T, and T47D) with the intent of further refining the chemicals' structural attributes related to cytostatic and cytotoxic activity to one type of breast cancer cell line. The results of these models are discussed below. Furthermore, since it became apparent that the modeling method was able to discern structural differences relating to differential potency of agents between breast cancer cell lines, this aim was further extended to cover several other tumor sites (e.g., breast vs. central nervous system, leukemia, melanoma, prostate, colon, lung, ovary, and kidney). Essentially, the preliminary results from these later models clearly suggest that structural differences exist between agents that are toxic to breast cancer cell and innocuous to cancer cells to other tissues of the body. The outcome of these results thus support one of the overall goals of this project wherein structural features that bestow toxicity toward breast cancer cells and minimal toxicity to other systems of the body may facilitate the development of antibreast cancer agents that are effective specifically against breast cancer cells and have minimal (or no) adverse effects on other cells of the body.

All of the models described above have been developed with the cat-SAR expert system. These include MCF-7, MDA-MB-213 models for GI_{50}, TGI, and $LC_{50,}$ and the project's key models, the excessive toxicity models (i.e., MCF-7 – MDA-MB-213 (MCF-MDA) and MDA-MB-231 – MCF- 7 (MDA-MCF)). A total of 12 base-line (see Table 1) and final rule optimized (see Table 2) models were developed with each model consisting of 400 compounds selected from the NCI's Developmental Therapeutics Program (DTP) with the desired activity (i.e., potency for the MCF-7 and MDA-MB-231 models and excessive toxicity to only one cell line for the MCF-MDA and MDA-MCF models). Likewise models including the breast cancer cell lines BT549, HS578T, and T47D have also been developed including differential activity models for MCF-7 and all other breast cancer lines (see Table 3 and 4). These models have been developed for the GI_{50} endpoint. And finally, GI_{50} and differential toxicity models for MCF-7 cells against several other tumor sites, including SNB-19 central nervous system, SR leukemia, M14 melanoma, prostate, HT29 colon, lung, IGROV1 ovary, EKVX lung, and A498 renal, have been successfully developed (see Tables 5 and 6).

First, a series of range-finding experiments were carried out in order to select the best overall modeling parameters. These included 2-D fragment lengths and cat-SAR modeling parameters (e.g., number of chemicals and proportions of active and inactive compounds required to select important fragments). Fragment length of three to seven, and ones with single fragment lengths from seven to 12 heavy atoms were analyzed. Originally, we selected the eight heavy atom fragment size models for further work since 1) models built on them were able to predict a significant number of compounds from the learning sets and eight heavy atoms seemed an appropriate size for developing pharmacophores for later 3-D QSAR modeling. Moreover, from a practical point, models of atom size 12 had roughly nearly 200,000 fragments wherein size eight had about 84000. This reduction in fragments made data management and analysis more practical without losing predictive ability. However, it was ultimately determined that the models of fragment size 3-7 were more applicable for this study. This was based on the necessity to adequately compare models across endpoints (i.e., most of the other models we have developed for cancer, mutagenesis, and mammary cancer related endpoints recently are based on fragment length of 3-7).

Additionally, as part of this study, a new method was developed to optimize cat-SAR models. Previous cat-SAR models required a relatively arbitrary setting of parameters that selected important fragments (fragment compound counts and fragment activity proportion values). For these analyses a rule optimization routine was employed. The optimization routine in this instance allowed the Number Rule to range between 1 and 8 and the Proportion Rule to range between 0.50 and 0.95. Leave-one out (LOO) validations were then conducted for each model and final models were selected that were both highly accurate (i.e., had a high concordance between experimental and predicted values) and highly predictive (i.e., made predictions on most of the chemicals in the learning set).

Table 1. Baseline summary, self-fit, and cross validation results for MCF-7, MDA-MB-231, and differential activity models

					Self-fit			LOO			LMO		
Model	**Total**	**Model**	**Active**	**Inactive**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**
GI50													
MCF-7	49958	10135	7673	2462	0.90 (180/200)	0.86 (170/197)	0.88 (350/397)	0.80 (159/200)	0.79 (156/198)	0.79 (315/398)	0.81 (4.0/5.0)	0.79 (3.9/4.9)	0.80 (7.9/9.8)
MDA-MB-231	58821	10006	6718	3288	0.89 (177/200)	0.89 (176/197)	0.89 (353/397)	0.85 (170/200)	0.78 (154/197)	0.82 (324/397)	0.86 (4.3/5.0)	0.69 (3.3/4.8)	0.77 (7.6/9.8)
MCF-7 Diff	52316	6661	3538	3123	0.84 (165/196)	0.84 (164/196)	0.84 (329/392)	0.74 (146/198)	0.68 (132/194)	0.71 (278/392)	0.81 (4.0/5.0)	0.51 (2.4/4.7)	0.66 (6.4/9.7)
MDA-MB-231 Diff	57637	5929	2875	3054	0.81 (157/193)	0.82 (163/200)	0.81 (320/393)	0.58 (111/192)	0.58 (115/198)	0.58 (226/391)	0.60 (2.8/4.7)	0.60 (3.0/5.0)	0.60 (5.8/9.7)
TGI													
MCF-7	50013	10095	7093	3002	0.91 (182/200)	0.85 (167/197)	0.88 (349/397)	0.72 (144/200)	0.74 (146/198)	0.73 (290/398)	0.72 (3.5/4.8)	0.75 (3.7/5.0)	0.73 (7.2/9.8)
MDA-MB-231	56716	9859	6613	3246	0.86 (172/200)	0.89 (176/198)	0.87 (348/398)	0.82 (164/200)	0.80 (158/198)	0.81 (322/398)	0.82 (4.1/5.0)	0.81 (3.9/4.8)	0.81 (8.0/9.8)
MCF-7 Diff	53837	7616	4373	3243	0.80 (160/200)	0.80 (156/196)	0.80 (316/396)	0.70 (139/199)	0.69 (136/196)	0.70 (275/395)	0.68 (3.3/4.9)	0.70 (3.4/4.9)	0.69 (6.7/9.8)
MDA-MB-231 Diff	51267	7215	4665	2550	0.83 (163/197)	0.83 (163/196)	0.83 (326/393)	0.68 (133/169)	0.68 (132/195)	0.68 (265/391)	0.67 (3.2/4.9)	0.69 (3.2/4.8)	0.67 (6.5/9.7)
LC50													
MCF-7	48699	8794	5859	2935	0.82 (165/199)	0.84 (162/192)	0.84 (327/391)	0.69 (137/199)	0.71 (134/190)	0.70 (271/389)	0.71 (3.4/4.9)	0.70 (3.3/4.7)	0.70 (6.7/9.6)
MDA-MB-231	57674	9489	6302	3187	0.84 (167/199)	0.82 (163/198)	0.83 (330/397)	0.76 (152/200)	0.74 (146/198)	0.75 (298/398)	0.77 (3.8/4.9)	0.75 (3.6/4.9)	0.75 (7.4/9.8)
MCF-7 Diff	55053	7636	3742	3894	0.81 (159/197)	0.81 (160/198)	0.81 (319/395)	0.70 (137/196)	0.69 (137/198)	0.70 (274/394)	0.70 (3.4/4.9)	0.70 (3.3/4.8)	0.69 (6.7/9.7)
MDA-MB-231 Diff	51300	7128	3539	3589	0.80 (160/199)	0.80 (158/197)	0.80 (318/396)	0.68 (135/198)	0.69 (135/196)	0.69 (270/394)	0.69 (3.4/4.9)	0.69 (3.3/4.8)	0.69 (6.7/9.7)

Notes:
Total Fragments: number of fragments derived from learning set.
Model Fragments: number of fragments meeting specified rules of the model.
Active Fragments: number of fragments meeting specified rules to be considered as active.
Inactive Fragments: number of fragments meeting specified rules to be considered as inactive.
Sensitivity: number of correct positive predictions / total number of positive predictions.
Specificity: number of correct negative predictions / total number of negative predictions
Concordance: Observed Correct Predictions: number of correct predictions / total number of predictions.

Table 2. Parameter optimized fragment summary, self-fit, and cross validation results for MCF-7, MDA-MB-231, and differential activity models.

					Self-fit			LOO			LMO		
Model	**Total**	**Model**	**Active**	**Inactive**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**
GI50													
MCF-7	49958	40305	18982	21323	0.95 (190/200)	0.95 (184/193)	0.95 (374/393)	0.83 (165/200)	0.83 (160/194)	0.83 (325/394)	0.83 (4.0/4.9)	0.83 (3.9/4.7)	0.83 (8.0/9.7)
MDA-MB-231	58821	51329	23809	27520	0.94 (187/200)	0.96 (191/199)	0.95 (378/399)	0.84 (168/200)	0.85 (169/199)	0.85 (337/399)	0.84 (4.1/4.9)	0.83 (4.0/4.8)	0.84 (8.2/9.8)
MCF-7 Diff	52316	1637	1519	118	0.80 (131/164)	0.80 (125/157)	0.80 (256/321)	0.79 (125/159)	0.79 (130/167)	0.78 (255/326)	0.75 (3.0/4.0)	0.76 (3.1/4.0)	0.76 (6.1/8.0)
MDA-MB-231 Diff	57637	217	123	94	0.89 (110/85)	0.81 (69/85)	0.86 (179/208)	0.7 0(83/119)	0.68 (70/103)	0.69 (153/222)	0.71 (2.0/2.9)	0.67 (1.6/2.5)	0.69 (3.7/5.5)
TGI													
MCF-7	50013	44033	25192	18841	0.90 (179/200)	0.92 (183/199)	0.91 (362/399)	0.84 (167/200)	0.79 (157/199)	0.81 (324/399)	0.82 (3.9/4.8)	0.79 (3.9/5.0)	0.81 (7.9/9.8)
MDA-MB-231	56716	47656	17615	30041	0.96 (192/200)	0.97 (194/200)	0.97 (386/400)	0.87 (173/200)	0.83 (164/197)	0.85 (337/397)	0.86 (4.3/5.0)	0.84 (4.0/4.7)	0.85 (8.3/9.9)
MCF-7 Diff	53837	5370	4563	807	0.83 (165/199)	0.84 (163/194)	0.84 (328/393)	0.76 (152/199)	0.73 (141/194)	0.75 (293/393)	0.71 (3.4/4.8)	0.73 (3.5/4.8)	0.72 (7.0/9.7)
MDA-MB-231 Diff	51267	664	433	231	0.88 (133/151)	0.85 (106/121)	0.88 (239/272)	0.76 (111/147)	0.76 (103/136)	0.76 (214/283)	0.74 (2.7/3.7)	0.76 (2.5/3.3)	0.75 (5.2/7.0)
LC50													
MCF-7	48699	2413	1872	541	0.80 (159/198)	0.79 (144/185)	0.79 (303/383)	0.76 (152/199)	0.75 (142/189)	0.76 (294/388)	0.75 (3.6/4.8)	0.75 (3.4/4.6)	0.75 (7.1/9.5)
MDA-MB-231	57674	17977	12315	5662	0.89 (178/200)	0.85 (170/199)	0.87 (348/399)	0.84 (167/200)	0.79 (158/199)	0.82 (325/399)	0.81 (4.0/4.9)	0.81 (3.9/4.9)	0.81 (7.9/9.8)
MCF-7 Diff	55053	1325	651	674	0.88 (119/136)	0.88 (132/150)	0.88 (252/286)	0.76 (102/135)	0.75 (116/154)	0.75 (218/289)	0.75 (2.5/3.4)	0.77 (2.8/3.7)	0.75 (5.3/7.1)
MDA-MB-231 Diff	51300	3089	1998	1091	0.90 (165/184)	0.91 (159/184)	0.91 (324/359)	0.76 (140/184)	0.76 (130/171)	0.76 (270/355)	0.76 (3.4/4.5)	0.75 (3.1/4.2)	0.75 (6.5/8.7)

Notes: see Table 1

Table 3. Baseline fragment summary, self-fit, and cross validation results for additional breast cancer cell lines and differential activity models.

					Self-fit			LOO			LMO		
Model	**Total**	**Model**	**Active**	**Inactive**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**
GI50													
MCF-7	49958	10135	7673	2462	0.90 (180/200)	0.86 (170/197)	0.88 (350/397)	0.80 (159/200)	0.79 (156/198)	0.79 (315/398)	0.81 (4.0/5.0)	0.79 (3.9/4.9)	0.80 (7.9/9.8)
MDA-MB-231	58821	10006	6718	3288	0.89 (177/200)	0.89 (176/197)	0.89 (353/397)	0.85 (170/200)	0.78 (154/197)	0.82 (324/397)	0.86 (4.3/5.0)	0.69 (3.3/4.8)	0.77 (7.6/9.8)
BT549								0.79 (157/200)	0.76 (152/199)	0.78 (309/399)			
HS578T								0.80 (159/200)	0.82 (163/199)	0.81 (322/399)			
T47D								0.80 (160/199)	0.83 (163/197)	0.82 (323/396)			
MCF-7 MDA-MB-231	52316	6661	3538	3123	0.84 (165/196)	0.84 (164/196)	0.84 (329/392)	0.74 (146/198)	0.68 (132/194)	0.71 (278/392)	0.81 (4.0/5.0)	0.51 (2.4/4.7)	0.66 (6.4/9.7)
MCF-BT								0.75 (150/200)	0.75 (146/196)	0.75 (296/396)			
MCF-HS								0.76 (151/200)	0.75 (148/198)	0.75 (299/398)			
MCF-T4								0.70 (139/198)	0.69 (138/199)	0.70 (277/397)			
MCF-MDA-T4								0.67 (133/200)	0.68 (134/197)	0.67 (267/397)			
MCF-MDA-T47D-HS578T								0.71 (142/200)	0.71 (140/198)	0.71 (282/398)			
MCF-MDA-T4-HS578T-BT549								0.69 (137/199)	0.68 (136/200)	0.68 (273/399)			

Notes: see Table 1

Table 4. Parameter optimized fragment summary, self-fit, and cross validation results for additional breast cancer cell lines and differential activity models.

					Self-fit			LOO			LMO		
Model	**Total**	**Model**	**Active**	**Inactive**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**
GI50													
MCF-7	49958	40305	18982	21323	0.95 (190/200)	0.95 (184/193)	0.95 (374/393)	0.83 (165/200)	0.83 (160/194)	0.83 (325/394)	0.83 (4.0/4.9)	0.83 (3.9/4.7)	0.83 (8.0/9.7)
MDA	58821	51329	23809	27520	0.94 (187/200)	0.96 (191/199)	0.95 (378/399)	0.84 (168/200)	0.85 (169/199)	0.85 (337/399)	0.84 (4.1/4.9)	0.83 (4.0/4.8)	0.84 (8.2/9.8)
BT549								0.81 (159/197)	0.84 (164/196)	0.82 (323/393)			
HS578T								0.84 (168/199)	0.84 (163/194)	0.84 (331/393)			
T47D								0.85 (169/200)	0.86 (172/199)	0.86 (341/399)			
MCF-MDA	52316	1827	1698	129	0.81 (139/172)	0.79 (126/159)	0.80 (265/331)	0.78 (125/160)	0.75 (129/172)	0.77 (254/332)	0.75 (3.0/4.0)	0.74 (3.0/4.1)	0.75 (6.1/8.2)
MCF-BT								0.77 (154/200)	0.80 (157/197)	0.78 (311/397)			
MCF-HS								0.80 (159/200)	0.81 (162/199)	0.81 (321/399)			
MCF-T47D								0.71 (142/200)	0.78 (155/200)	0.74 (297/400)			
MCF-MDA-T47D								0.74 (148/200)	0.77 (153/200)	0.75 (301/400)			
MCF-MDA-T47D-HS								0.74 (148/200)	0.73 (146/200)	0.74 (294/400)			
MCF-MDA-T4-HS-BT								0.74 (147/199)	0.74 (146/198)	0.74 (293/397)			

Notes: see Table 1

Table 5. Baseline fragment summary, self-fit, and cross validation results for breast cancer and other target tissue models.

					Self-fit			LOO			LMO		
Model	**Total**	**Model**	**Active**	**Inactive**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**
GI50													
MCF-7	49958	10135	7673	2462	0.90 (180/200)	0.86 (170/197)	0.88 (350/397)	0.80 (159/200)	0.79 (156/198)	0.79 (315/398)	0.81 (4.0/5.0)	0.79 (3.9/4.9)	0.80 (7.9/9.8)
CNS								0.82 (164/200)	0.80 (158/198)	0.81 (322/398)			
COLON								0.82 (163/200)	0.83 (163/196)	0.82 (326/396)			
LEUK								0.83 (166/200)	0.86 (167/195)	0.84 (333/395)			
LUNG								0.80 (160/200)	0.81 (160/197)	0.81 (320/397)			
MELANOMA								0.79 (157/200)	0.78 (155/199)	0.78 (312/399)			
OVARY								0.85 (170/200)	0.86 (166/194)	0.85 (336/394)			
PROSTATE								0.81 (161/200)	0.82 (160/195)	0.81 (321/395)			
RENAL								0.85 (169/200)	0.85 (164/193)	0.85 (333/393)			
MCF7-CNS								0.68 (135/199)	0.68 (134/197)	0.68 (269/396)			
MCF-LEUK								0.60 (118/197)	0.61 (120/197)	0.60 (238/394)			
MCF-MELANO								0.72 (143/198)	0.72 (144/200)	0.72 (287/398)			
MCF-PROST								0.68 (136/200)	0.68 (134/197)	0.68 (270/397)			
MCF-COLON								0.75 (149/199)	0.74 (147/199)	0.74 (296/398)			
MCF-LUNG								0.72 (143/200)	0.71 (139/195)	0.71 (282/395)			
MCF-OVARY								0.71 (140/198)	0.71 (139/197)	0.71 (279/395)			
MCF-RENAL								0.67 (131/197)	0.67 (130/195)	0.67 (261/392)			
MCF-LUNG-RENAL								0.68 (134/198)	0.67 (132/197)	0.67 (266/395)			
MCF-LUNG-RENAL-CNS								0.70 (136/194)	0.70 (138/197)	0.70 (274/391)			
MCF-LUNG-RENAL-COLON-LEUK								0.64 (127/198)	0.64 (125/195)	0.64 (252/393)			
MCF-CNS-COLON-LUNG-RENAL-LEUKEMIA								0.65 (125/193)	0.64 (127/198)	0.65 (252/391)			
MCF-CNS-COLON-KID-LEUK-LUNG-MEL								0.59 (114/192)	0.60 (107/179)	0.60 (221/371)			

Notes: see Table 1

Table 6. Parameter optimized fragment summary, self-fit, and cross validation results for breast cancer and other target tissue models.

					Self-fit			LOO			LMO		
Model	**Total**	**Model**	**Active**	**Inactive**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**	**Sensitivity**	**Specificity**	**Concordance**
GI50													
MCF-7	49958	40305	18982	21323	0.95 (190/200)	0.95 (184/193)	0.95 (374/393)	0.83 (165/200)	0.83 (160/194)	0.83 (325/394)	0.83 (4.0/4.9)	0.83 (3.9/4.7)	0.83 (8.0/9.7)
CNS								0.87 (173/200)	0.81 (161/199)	0.84 (334/399)			
COLON								0.85 (169/200)	0.82 (156/190)	0.83 (325/390)			
LEUK								0.88 (176/200)	0.85 (166/196)	0.86 (342/396)			
LUNG								0.86 (172/200)	0.84 (167/198)	0.85 (339/398)			
MELANOMA								0.85 (170/200)	0.84 (166/197)	0.85 (336/397)			
OVARY								0.86 (171/200)	0.88 (173/196)	0.87 (344/396)			
PROSTATE								0.84 (169/200)	0.85 (165/195)	0.85 (334/395)			
RENAL								0.86 (172/200)	0.88 (172/195)	0.87 (344/395)			
MCF7-CNS								0.73 (141/193)	0.74 (139/189)	0.73 (280/382)			
MCF-LEUK								0.64 (127/200)	0.50 (25/50)	0.66 (264/399)			
MCF-MELANO								0.75 (149/199)	0.76 (152/200)	0.75 (301/399)			
MCF-PROST								0.73 (146/200)	0.75 (148/197)	0.74 (294/397)			
MCF-COLON								0.77 (152/198)	0.76 (149/196)	0.76(301/394)			
MCF-LUNG								0.78 (155/198)	0.76 (147/193)	0.77 (302/391)			
MCF-OVARY								0.72 (144/200)	0.74 (147/198))	0.73 (291/398)			
MCF-RENAL								0.75 (149/198)	0.75 (145/194)	0.75 (294/392)			
MCF-LUNG-RENAL								0.70 (137/195)	0.71 (139/196)	0.71 (276/391)			
MCF-LUNG-RENAL-CNS								0.81 (161/200)	0.75 (150/200)	0.78 (311/400)			
MCF-LUNG-RENAL-COLON-LEUK								0.70 (136/195)	0.70 (138/197)	0.70 (274/392)			
MCF-CNS-COLON-LUNG-RENAL-LEUKEMIA								0.65 (128/196)	0.74 (148/199)	0.70 (276/395)			
MCF-CNS-COLON-KID-LEUK-LUNG-MEL								0.68 (132/194)	0.69 (125/182)	0.68 (257/376)			

Notes: see Table 1

Model Analysis and Validation

1. MCF-7 and MDB-MB-231 cell lines

Together, 24 cat-SAR models were produced. The first set of models were developed with a standard values for selecting important cat-SAR fragments. For these models the Number Rule was set to three and the Proportion Rule was set to 0.75 for both the active and inactive category (Table 1). The second and final set of models was then developed with the cat-SAR Rule Optimizer that explored values for the Number and Proportion Rules in order to develop models with best concordance between experimental and predicted values (Table 2). This same methodology applies to models and results discussed below and in Tables 3-6.

Overall, between 48,699 and 58,821 unique chemical fragments between 3-7 non-hydrogen atoms were derived for these models (Tables 1 and 2). The self-fit analysis yielded concordance between observed and predicted results of 80 and 89% for the standard or base line models (Table 1) and between 80 and 97% for the rule optimized ones (Table 2). These high concordance rates across all models indicate that the models are robust and there is sufficient structural information contained in the learning sets to distinguish between active and inactive compounds.

The LOO cross-validation concordance values ranged between 58 and 82% for the standard models and between 69 and 85% for the rule optimized ones. As expected, the model derived for single cell lines generally had a higher concordance value than those for the differential activity models that compared cell lines. For example, the MCF-7 and MDA-MB-231 GI50 models returned concordance values of 83 and 85%, respectively, while the MCF-MDA and MDA-MCF models had concordance values of 77 and 69%, respectively.

In all cases, the LMO cross-validations nearly equaled the LOO validation results verifying that there is sufficient structural information contained in the learning sets to distinguish between active and inactive compounds.

In order to better judge how well these two models performed in general, one can consider the “accuracy” or reproducibility of a standard in vitro toxicological test. For instance, the US National Toxicology Program’s (NTP) Salmonella mutagenicity database, which is derived from a standardized protocol, has been estimated to be about 85% reproducible (49). Based on these findings wherein the concordance between observed and predicted values for human developmental toxicity is ~80% the cat-SAR human developmental toxicity models thus appear to be as predictive as data used to develop SAR models from standardized in vitro assays.

The overall results obtained from this exercise show that cat-SAR expert system can make mechanistically sound models that are capable of generating information which can be used to identify the chemical moieties specific to MCF-7 cell line. We have observed that the cat-SAR expert system produces models which are predictive and are based on mechanically sound attributes.

2. Extension to other breast cancer cell lines

Tables 3 and 4 cover LOO validation results for models we extended from MCF-7 and MDA-MB-231 cell lines to three other breast cancer cell lines (i.e., BT549, HS578T, and T47D). Although analyses of these data as of June 2009 are just commencing, it is clear that data from BT549, HS578, and T47D cell lines are capable of yielding SAR models with concordances between experimental and predicted results between 84 and 86%. Furthermore, looking at the differential activity models between MCF-7 and the other breast cancer cell lines simultaneously, the concordance value is 74%. This last value suggests that structural information can be identified that is associated between toxicity to one cell line (MCF-7) and not to four other breast cancer cell lines. It should be noted at this juncture that additional model creation and validations are needed to fill-out Tables 3 and 4 including possibly looking at TGI and LC_{50} endpoints.

3. Extension to other non-breast cancer cell lines

Tables 5 and 6 cover the LOO validation results for MCF-7 and eight other target tissues, along with differential activity models for MCF-7 and up to 8 different tissues simultaneously. Of the eight individual models, all had concordance values range between 83 and 87%. Considering the MCF-7 differential activity models for the different cell lines, concordance values were mostly in the 70%s. Importantly, these values held in the 70%s even for differential activity models built on multiple cell lines. Overall, these preliminary results strongly suggest that there are structural differences in chemicals the bestow toxicity to MCF-7 breast cancer cells and are not toxic against other tissue types. It should be noted again at this point that additional model creation and validations are needed to fill-out Tables 5 and 6 including possibly looking at TGI and LC_{50} endpoints. Likewise, we anticipate producing several models that generally compare breast cancer cell lines (rather than just MCF-7) to other tissue types.

Specific Aim 2:

As discussed in the previous report, we explored and developed CoMFA models of several fragments identified in the MCF-MDA and MDA-MCF GI_{50} models. After these models were completed we passed along some recommended compounds for proteomic testing to our collaborator Dr. Day at the University of Pittsburgh. Currently Dr. Day is working on obtaining test chemicals from the NCI.

Specific Aim 3:

This specific aim as described in the statement of work will entail differential proteomics analysis between breast cancer cell lines that respond differentially to cytostatic and cytotoxic agents. At this time Dr. Day is still working on obtaining compounds from the NCI. In the mean time, Dr. Day has been determining differentially expressed proteins in MCF-7 and MCF-7/LY2 cells exposed to tamoxifen. Essentially, Dr. Day has been focusing on one compound, tamoxifen, and has been readying the experimental processes required for Specific Aim 3 once more compounds are obtained from the NCI

With regard to Specific Aim 3, the following is an abstract from Miranda Sarachine's Ph.D. dissertation.

Tamoxifen has long been the standard of treatment in women with ER+ breast cancer. One of the major problems with tamoxifen is drug resistance. Only about half of ER+ breast cancers respond to tamoxifen initially, and those that do respond eventually develop resistance over the course of treatment. The mechanism behind this resistance is not completely understood. Several mechanisms have been proposed, such as alterations to cofactors, loss of ER expression, and compensation by growth factor signaling pathways. In order to investigate the alterations to the proteins of the nuclear matrix (NM), the NM proteins (NMPs) from MCF-7 breast cancer cells, and their antiestrogen resistance derivative, MCF-7/LY2, were isolated and examined. These two cell lines were also exposed to several ER isoform-selective ligands in order to discern how the receptor isoforms may alter NMP composition of tamoxifen-responsive and -resistant cell lines. The 8-plex iTRAQ method was used to identify and quantify the NMPs present in these cell lines. A total of 148 NMPs were identified, and several interesting changes between the resistant and responsive lines, and in response to the subtype selective ligands, were found. These proteins may shed some light on the mechanism behind antiestrogen resistance and eventually serve as biomarkers to help customize breast cancer treatment.

Furthermore, in the event that we cannot obtain compounds from the NCI to address specific aim 3, we have devised two backup plans. First, we plan to analyze results of compounds Dr. Day already tested and published to determine if they fit into the results obtained from our modeling experiments. Second, as part of another project, we have developed an SAR modeling approach that has been successful at identify specific protein target that are capable of distinguishing rat mammary carcinogens from other carcinogens to other tissues. Hence we have hypothesized that these models are essentially identifying unique molecular targets associated with chemically induced breast cancer. To this end, I have secured the computational resources and the collaborative effort of Professor John Trent (University of Louisville) to virtually screen some of the reported breast cancer and excessive toxicity models against a set of ~5500 protein targets. The end results of this in silico approach will essentially be the same as the projects in vitro goal of identifying specific molecular targets that are associated with cytostatic and cytoxic activity specifically against breast cancer cells.

Key Research Accomplishments

As described above

- SAR models have been produced and validated for MCF-7 and MDA-MB-231 GI_{50}, TGI, and LC_{50} individually and for differential activity when comparing MCF-7 and MDA-MB-231 cells. Hence, SA1 is completed.
- The first CoMFA/CoMSIA models for SA2 have been preliminarily analyzed. Results suggest that the proposed use of CoMFA/CoMSIA 3-dimensional modeling will be useful in the decision making process for selecting the most promising compounds for proteomic analyses in SA3.
- The tools and techniques to accomplish SA3 have been worked over in the Day lab with all indications that proteins differentially expressed in breast cancer cell lines exposed to the same chemicals will be realized.

Reportable Outcomes

1. Sarachine, M.J., Janjic, J.M., Wipf, P., Day, B.W. Biphenyl C-cyclopropylalkylamides: New scaffolds for targeting estrogen receptor beta (J. Bioorganic & Medicinal Chemistry Letters 2009, 19, 2404-2408.
2. Sarachine, M.J., Janjic, J.M., Wipf, P., Day, B.W. Biphenyl C-cyclopropylalkylamides as new scaffolds for targeting estrogen receptor beta (J. Proceedings of the 2008 Era of Hope Department of Defense Breast Cancer Research Program Meeting Baltimore, MD, June 2008, p. 104 (P15-11).
3. Sarachine, M.J., Latimer, J.L., Day, B.W. Quantitative proteomics of nuclear matrix proteins in novel human ductal carcinoma in situ model systems. Proceedings of the 2008 Era of Hope Department of Defense Breast Cancer Research Program Meeting Baltimore, MD, June 2008, p. 178 (P26-8).
4. Sarachine, M.J., Janjic, J.M., Wipf, P., Day, B.W. Biphenyl C-cyclopropylalkylamides as new scaffolds for targeting estrogen receptor beta. Proceedings of the 31st National Medicinal Chemistry Symposium Vol. 31, Pittsburgh, PA, June 2008, p. 88 (233).
5. Qamar, S., C.A. Carrasquer, S.L. Cunningham, and A.R. Cunningham. Structure-Activity Relationship Model for Differential Growth Inhibition of MCF-7 and MDA-MB-231 Cells, Society of Toxicology, Baltimore, MD, March 15-19, 2009.
6. Carrasquer, C.A., S. Qamar, and A.R. Cunningham. Structure-Activity Relationship Analyses and the Identification of Mammary Anticarcinogens, Society of Toxicology, Baltimore, MD, March 15-19, 2009.
7. Manuscript in preparation: Shahid Qamar, Carl A. Carrasquer, and Albert R. Cunningham. Structure-Activity Relationship Model for Differential Growth Inhibition of MCF-7 and MDA-MB-231 Cells.
8. Manuscript in preparation: Shahid Qamar, Carl A. Carrasquer, and Albert R. Cunningham. Structure-Activity Relationship Model for Differential Growth Inhibition between human breast cancer cells and other tumor target sites.
9. Manuscript in preparation: Sarachine, M.J., Hall, C., Cunningham, A.R, Day, B.W. Alterations to the nuclear matrix in antiestrogen resistance and the impact of estrogen receptor selective ligands.
10. Manuscript in preparation: Miranda J. Sarachine, Tamanna Sultantana, Mirunalni Thangavelu, Manimalha Balasubramani, Albert R. Cunningham, and Billy W. Day. Identification and Quantitation of Differentially Expressed Proteins in Tamoxifen-Resistant Breast Cancer Cells.
11. The list of SAR databases outlined in Tables 1 and 2 have been produced and validated.
12. Miranda J. Sarachine, was a School of Medicine graduate student working in the Day Lab at the University of Pittsburgh and was being partially funded through this grant's subcontract with the Unversity of Pittsburgh, and we are pleased to announce that Miranda has been awarded her Ph.D.

Conclusion

As of July 2009, with the series of well-working models developed in SA1 that are capable of analyzing not only cytostatic and cytotoxic activity to MCF-7 and MDA-MB-231 cells but also describing differences in activities induced by small molecules to these cell lines, as well as the models extended to several other breast cancer cell lines and even other tissues, we are confident that this project will achieve the goal set out in the proposal, i.e., the identification of molecular targets and pharmacophores that are exquisitely associated with high toxicity and breast caner cell specificity. Furthermore, with the success of the preliminary CoMFA and CoMSIA modeling described in the previous report, we are confident that we will be able to fine-tune the selection process of compounds for proteomics analysis. In essence, the initial success of SA2 indicates that this 3-dimenstional QSAR modeling approach will allow for a refined set of compounds for testing in SA 3

And finally, with the initial work conducted by Dr. Day with regards to protein identification in different breast cancer cell lines exposed to the same compounds, we are confident that once the appropriate compounds are obtained from the NCI, SA3 will not encounter any significant pitfalls.

References

None